Cognitive Science 33 (2009) 1187–1191

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01064.x

Sequence Encoders Enable Large-Scale Lexical Modeling: Reply to Bowers and Davis (2009)

Daragh E. Sibley,[a] Christopher T. Kello,[b] David C. Plaut,[c] Jeffrey L. Elman[d]

[a]*Department of Psychology, George Mason University*
[b]*Cognitive Science Program, University of California, Merced*
[c]*Department of Psychology, Center for the Neural Basis of Cognition, Carnegie Mellon University*
[d]*Department of Cognitive Science, University of California, San Diego*

Received 9 December 2008; received in revised form 26 March 2009; accepted 28 March 2009

Abstract

Sibley, Kello, Plaut, and Elman (2008) proposed the sequence encoder as a model that learns fixed-width distributed representations of variable-length sequences. In doing so, the sequence encoder overcomes problems that have restricted models of word reading and recognition to processing only monosyllabic words. Bowers and Davis (2009) recently claimed that the sequence encoder does not actually overcome the relevant problems, and hence it is not a useful component of large-scale word-reading models. In this reply, it is noted that the sequence encoder has facilitated the creation of large-scale word-reading models. The reasons for this success are explained and stand as counterarguments to claims made by Bowers and Davis.

Keywords: Large-scale modeling; Sequence encoder; Orthography; Phonology; Wordforms

We recently presented a connectionist architecture termed the ''sequence encoder'' that learns orthographic and phonological representations of words (Sibley, Kello, Plaut, & Elman, 2008). These models extend simple recurrent networks to learn distributed codes that represent elements of a sequence and their positions. Our goal in developing the sequence encoder was to learn representations shaped by graphotactic and phonotactic structure in large-scale wordform lexicons. We reported simulations that learned representations for over 70,000 English wordforms, ranging from 1 to 18 letters or phonemes in length, and 1 to 5 syllables in length. These sequence encoders accurately encoded untrained strings of letters and phonemes (i.e., pseudowords) that coincide with the graphotactics or phonotactics of the trained language. They also accounted for participants' judgments about the conformity of pseudowords to graphotactic regularities.

Correspondence should be sent to Daragh E. Sibley, Department of Psychology, George Mason University, Fairfax, VA 22030. E-mail: dsibley@wisc.edu

Until recently, nearly all models of word reading were restricted to monosyllabic words. This has often been attributed to the alignment and dispersion problems (Davis, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996): The alignment of letters or phonemes is ambiguous across wordforms of different lengths, and information about letters and phonemes is dispersed across position-specific units. These problems arise because previous models of word reading used slot-based representations, so that elements in different positions of a word were treated largely independently. The sequence encoder enables large-scale models of naming by converting variable-length sequences of letter or phonemes into distributed representations. Rueckl, Fang, Begosh, Rimzhim, and Tobin (2008) proposed a conceptually similar approach, while Ans, Carbonnel, and Valdois (1998) suggest a different solution.

Bowers and Davis (2009) criticized the sequence encoder in two respects. First, they argued that a system for representing wordforms must distinguishing words from nonwords, while the sequence encoder only distinguishes well-formed from ill-formed strings of letters or phonemes. Second, they argue that the sequence encoder does not resolve the dispersion and alignment problems because learned representations of letters and phonemes are not identical across positions and wordforms of different lengths. In this reply, we explain that Bowers and Davis' criticisms stem from a miscommunication of scope and a difference in goals.

The sequence encoder was not designed to distinguish words from nonwords. Bowers and Davis appear to have a different interpretation of the model's scope, possibly because of confusion with the term ''wordform,'' and also because of one sentence in which we wrote that ''the sequence encoder is a model of lexical performance in its own right'' (p. 752). Lexical performance in this context referred to behavioral measures of wordform processing predicted by the model (i.e., ratings of wordform well-formedness), not an ability to distinguish words from nonwords.

As for ''wordform,'' our use of the term is analogous to the well-known ''visual wordform area'' (VWA). The VWA is most active in response to words, less responsive to pseudowords, and still less responsive to nonwords (like consonant strings; Binder, Medler, Westbury, Liebenthal, & Buchanan, 2006; Dehaene, Cohen, Sigman, & Vinckier, 2005). The sequence encoder shows the same ordering in terms of encoding accuracy because it learns about graphotactics or phonotactics, not because it learns a lexicon. Just as the VWA is hypothesized to provide orthographic wordform representations (word or nonword) to the reading system, a sequence encoder may provide wordform representations to large-scale models of word reading. This goal was explicitly stated in our original paper, and we have used sequence encoders in models of word reading and recognition (Kello, 2006; Sibley, 2008; Sibley, Kello, & Seidenberg, in press).

The goal of large-scale modeling is also relevant to Bowers and Davis' second criticism that the sequence encoder does not resolve the dispersion and alignment problems. This criticism arises from a difference in goals. Bowers and Davis argue that resolving the dispersion and alignment problems requires that a model be able to recognize substrings within wordforms (i.e., CAT in TOMCAT) and represent substrings equivalently across positions (i.e., total position independence).

It is certainly true that readers can recognize substrings within wordforms, and tasks can likely be designed to elicit position-independent behaviors with substrings. We agree that the alignment and dispersion problems have limited the development of models that explain such tasks and behaviors. This motivated our efforts to develop a system for learning orthographic and phonological representations that facilitate simulations of word-reading performance for large-scale corpora of mono- and multisyllabic words and nonwords. Therefore, for our purposes, these problems are resolved if sequence encoders enable simulations of word-reading performance for large-scale corpora.

Kello (2006) and Sibley (2008) reported models of word naming and lexical decision that mapped orthographic wordforms, learned via sequence encoding, onto their phonological counterparts for tens of thousands of monosyllabic and multisyllabic English words. In Simulation 2 of Sibley (2008) lexical decision and naming accuracies for words were 91.4% and 94.3% correct, respectively, which is consistent with participants' responses to these same items in the English Lexicon Project (Balota et al., 2007). This simulation accounted for 41.6% and 35.4% of the variance in lexical decision and naming latencies for roughly 28,000 mono- and multisyllabic English words, respectively. Pseudoword naming accuracy was only 50.9% correct, but in an updated model, naming accuracies reached 86.8% and 65% for mono- and bisyllabic pseudowords, respectively (Sibley et al., in press). Further these models correctly predict adult readers' sensitivity to variables that impact naming latencies, like frequency, length, orthographic neighborhood, phonological neighborhood, consistency, syllabic length, stress typicality, and several of these variables interactions. In sum, these simulation results stand as clear evidence that the dispersion and alignment problems were resolved, for our purposes.

As we explained in our original manuscript, large-scale modeling is enabled because sequence encoder representations exhibit a *graded* similarity for letters or phonemes occurring in different positions, rather than position independence. Bowers and Davis argue that the lack of position-independent coding means that the sequence encoder does not solve the alignment problem. We concede that according to their definition, the sequence encoder does not solve the alignment problem. However, our simulations stand as evidence that the alignment problem, as defined by Bowers and Davis, is not a problem that needs to be solved in order to account for a large amount of mono- and multisyllabic naming data.

In fact, position independence is not desirable for our purposes because reading performance is affected by positional dependencies. For example, long vowels are more common in the initial versus final syllables of English wordforms because initial syllables are more commonly stressed. It is by learning such dependencies that the sequence encoder is able to distinguish well-formed from ill-formed wordforms.

Bowers and Davis argue that, despite graded similarity, sequence elements are not represented with enough similarity across positions. They base their argument on principal component analyses we reported to illuminate how sequence encoders normalize variable-length sequences. However, our analyses were not intended for this purpose, and it is misleading to use them as such. The problem is that principal components were extracted for individual elements, not entire wordform representations, and that our analysis was applied to the artificial corpus of Simulation 1, not the learned English orthography of

Simulation 2. The latter point is important because representations will differ when trained on corpora of random letter strings compared with English wordforms.

Here we provide an analysis that more directly addresses the graded nature of representations as a function of position. Graded similarity is expressed by anticipation errors (e.g., anticipating the T in SPAT to produce STAT) and perseveration errors (e.g., perseverating the P in SPAT to produce SAP) that cross intervening positions. Such errors are common in tasks like word naming and lexical decision, and they cannot occur above chance in the sequence encoder unless letters and phonemes are represented similarly across positions. We analyzed errors made by the orthographic sequence encoder reported as Simulation 2 (11% error rate for words and 24% for pseudowords) and found that 76% of them contained one or more anticipation or perseveration errors. Of these errors, 35.8% crossed one position, 32.3% crossed two, 13.7% crossed three, 7.6% crossed four, 4.6% crossed five, and 2.9% crossed six (when positional distance was ambiguous, as in BALL to LALL, the shorter distance was counted). By comparison, the mean chance rate of making each error was 5.4% based on single-letter frequencies in the models' training corpus, and the average length of a word in which an error occurred was 10.1 letters.

In summary, we view the graded similarity of letters or phonemes across positions, and graded levels of accuracy in processing well-formed words versus pseudowords versus ill-formed nonwords, as positive attributes of the sequence encoder rather than as problems.

Acknowledgments

D. E. Sibley was supported by the training grant T32 HD49899 from the NICHD.

References

Ans, B., Carbonnel, S., & Valdois, S. (1998). A connectionist multiple-trace memory model for polysyllabic word Reading. *Psychological Review*, *105*, 678–723.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchinson, K. I., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445–459.

Binder, J. R., Medler, D. A., Westbury, C. F., Liebenthal, E., & Buchanan, L. (2006). Tuning of the human left fusiform gyrus to sublexical orthographic structure. *NeuroImage*, *33*, 739–748.

Bowers, J., & Davis, C. (2009). Learning representations of wordforms with recurrent networks: Comment on Sibley, Kello, Plaut, & Elman. *Cognitive Science*, *33*(7), 1183–1186.

Davis, C. J. (1999). *The self-organizing lexical acquisition and recognition (SOLAR) model of visual word recognition*. Unpublished doctoral dissertation. Australia: University of New South Wales.

Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words: A proposal. *Trends in Cognitive Science*, *9*(7), 335–341.

Kello, C. T. (2006). Considering the junction model of lexical processing. In S. Andrews (Eds.), *From Inkmarks to ideas: Current issues in lexical processing*. New York: Psychology Press.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*, 56–115.

Rueckl, J. G., Fang, S., Begosh, K. T., Rimzhim, A., & Tobin, S. J. (2008). *Learned internal representations and letter position information: A connectionist approach*. Presented at Psychonomics, 2008. Chicago, Illinois.

Sibley, D. E. (2008). *Large scale modeling of single word reading and recognition*. Unpublished PhD thesis. Fairfax, VA: George Mason University.

Sibley, D. E., Kello, C. T., Plaut, D. C., & Elman, J. L. (2008). Large-scale modeling of wordform learning and representations. *Cognitive Science*, *32*, 741–754.

Sibley, D. E., Kello, C. T., & Seidenberg, M. S. (In press). Learning orthographic and phonological representations in model of monosyllabic and bisyllabic naming. *European Journal of Cognitive Psychology*.